Toward Next Generation Coaching Tools for Court Based Racquet Sports

Damien Connaghan and Noel E. O'Connor
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Dublin 9
Ireland
connagha@eeng.dcu.ie

ABSTRACT

Even with today's advances in automatic indexing of multimedia content, existing coaching tools for court sports lack the ability to automatically index a competitive match into key events. This paper proposes an automatic event indexing and event retrieval system for tennis, which can be used to coach from beginners upwards. Event indexing is possible using either visual or inertial sensing, with the latter potentially providing system portability. To achieve maximum performance in event indexing, multi-sensor data integration is implemented, where data from both sensors is merged to automatically index key tennis events. A complete event retrieval system is also presented to allow coaches to build advanced queries which existing sports coaching solutions cannot facilitate without an inordinate amount of manual indexing.

Categories and Subject Descriptors

H.5.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing methods*

General Terms

Algorithms, Experimentation, Performance

Keywords

Sports Coaching, Event Indexing, Visual Sensing, Inertial Sensing

1. INTRODUCTION

This paper introduces a system which can automatically index all the main tennis events performed by both players in a match, with only one assumption, that a tennis match follow the rules of tennis as laid down by the International Tennis Federation[1]. The event indexing system can also infer which player has executed a specific stroke type in the match, which is a step forward for automatic event indexing in tennis. All events are uploaded to a database to facilitate a powerful event retrieval system.

[1]http://www.itftennis.com/technical/rules/

MM'12, October 29–November 2, 2012, Nara, Japan.

Approaches for visual event indexing for tennis have been previously reported [3] [6]. There has, however, been no content based event retrieval system which is facilitated by automatic stroke recognition in tennis or any other court based racquet sports such as badminton or pickleball . Issues with existing automatic sports coaching tools such as Hawke-Eye Coaching[2] include its high cost, its inability to automatically recognise which player has executed a stroke during a competitive match (due to change of ends) and more importantly its inability to automatically recognise the stroke type played, which is essential for enabling coaches to generate high-level tactical queries. Currently coaches need to spend endless hours manually editing existing coaching tools such as Dartfish[3]. This paper addresses all these problems to provide the first tennis event retrieval system which is driven by stroke recognition, along with player and ball localisation. Three individual components of this system were previously published in [1], but this paper makes several novel contributions. Firstly within the event indexing component, we introduce a novel method for detecting stroke sub categories and also what player performed a stroke. An additional contribution within the event retrieval system is the user query panel.

2. INFRASTRUCTURE

Visual Sensing: Three relatively low cost IP cameras are used, with pan, tilt and zoom (PTZ) capability. One camera provides an overhead view from the center of the court (AXIS 212, 13.8m height) and the other two baseline cameras are positioned at either end of the court (AXIS 215, 2.8m height). Additional cameras are optional.

Inertial Sensing: Each player wears a single Wireless Inertial Measuring Unit (WIMU) placed onto the player's dominant forearm during a competitive match. This provides a small, wearable and low cost method for instrumenting human subjects to provide high speed motion data using accelerometers, magnetometers and gyroscopes in each unit. Each WIMU connects wirelessly to a base station using a 2.4GHz RF system, which in turn connects to a court side PC [4].

3. EVENT INDEXING

To understand which automatically indexed events would benefit coaches in a retrieval system, we have worked closely with coaches to infer the system requirements. During this process the following events were found to be suitable for building comprehensive analysis queries: Serves (T, Body, Wide Serves); Forehands (in-to-in,

[2]http://www.hawkeyeinnovations.co.uk
[3]http://www.dartfish.com

in-to-out, cross, line); Backhands (cross, line); first serves made, first serves missed; return of first serves, return of second serves; net hits; rallies; games; player and ball localisation. Again, we stress that we, via our end user coaches, are motivated by automatic indexing for *coaching* rather than an extra competition umpire or visualisation system (such as Hawk-Eye).

To automatically index the required tennis events, the following three components were developed (1) Stroke Recognition, (2) Player and Ball Tracking, (3) Change of End Detection. *Stroke Recognition* is the process of identifying if the stroke played is a forehand, backhand, or serve. *Player and ball tracking* is the ability to track both players and the ball over the duration of the match. Robust *Change of End* detection is necessary to infer game boundaries and for player identification.

The next section describes how tennis events are indexed using visual sensors and the second section briefly describes the process used to infer events from inertial sensors.

3.1 Visual Event Indexing

This section describes the process used to automatically index events using video.

3.1.1 Player and Ball Tracking

Using the overhead camera, we have previously developed algorithms to detect the player and ball tracks [1]. The ball track gives the time and location of the origin of a ball hit and then the X,Y movement of the ball, along with the time until it stops. To detect both players from the aerial view camera, we use background subtraction and hysteresis-type blob tracking to track the tennis players positions.

3.1.2 Serve Detection

Using the data from the player tracks in Section 3.1.1, we can locate both player's positions and map these coordinates to the tennis court to determine each player's location on the court. Then, by determining both players' locations on the court at all times during a match, we are able to recognise a serve event when a player is located inside a serve zone for two seconds whilst their opponent is inside the diagonal return zone for five seconds and a ball hit occurs from the server's side of the court. This approach was previously published in [1].

3.1.3 Change of End Detection

Robust player identification and game identification is based on change of end times. Three steps are used to detect a Change of End, the first two are new and the third step has been previously published in [1]. *Step One* uses serve direction patterns (see Figure 1). Essentially this step uses the constraint from the International Tennis Federation rules that there is only one change of end (COE) event for every change of serve direction event (COSD). By serve direction, we mean the direction the ball travels after it is struck by the racquet, during a serve.

In *step two*, we need to find the best candidate for a change of end between two COSD events. We use player tracks to retrieve a temporal location where both players approach and/or walk from one side of the net to the other, as they will need to walk past the net at the change of ends. If this step finds more than one candidate change of end event (CCOE), between two COSD events, step three is required to check if there is a new player serving after a candidate change of end event.

Step Three exploits a recent fashion trend in tennis where players tend to wear colorful clothes. For every CCOE, the rear view camera is used to inspect the colour features of a serving player in

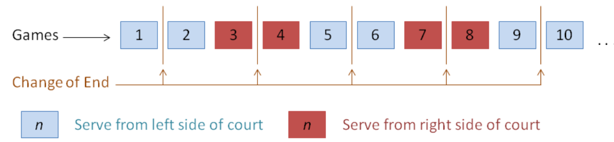

Figure 1: Pattern for serve direction and change of end, one change of end occurs for every change of serve direction

the previous serve and next serve. To identify if a new player is serving, the previous serve is flagged and 60 frames are extracted from the rear view camera. The player is extracted as a colour foreground from each image and the resulting HSV image is then split into three channels (Hue, Saturation and Value). We then create an image histogram from the Value channel only, which represents the brightness in the player foreground. We then apply an identical image processing technique to the next serve after the candidate change of end and the similarity distance between the histogram for the previous serve and the next serve is calculated using the Bhattacharyya coefficient, which has been widely used to compare image histograms [5]. A new player is deemed to be present in the image when the difference between the previous and next serve exceeds a threshold.

3.1.4 Backhand and Forehand Detection

We automatically remove the serves from the collection of ball hits. For each remaining ball hit in turn, we use the ball tracks (Section 3.1.1) for that stroke to find the origin of the stroke and the (x,y) position of the ball at the start of the stroke. We then extract the player as foreground from the video and obtain the centroid of the player as shown in Figure 3. This provides us with the position of the player and the position of the ball at the beginning of a stroke. Then we detect if the ball is to the left or right of the player at the start of the stroke and given that we can detect whether a player is right or left handed with the dominant arm detector (Section 3.1.5), we can infer if a stroke is a forehand or backhand.

3.1.5 Dominant Arm Detector

Using the camera behind the baseline as illustrated in Figure 2, we segment the player as the foreground for each image in a serve. Although the contour features that we use are robust to foreground holes and noise in the extracted silhouette, they can be adversely affected by shadows. For this reason, we use a layered background model that includes robust shadow detection. To extract contour features, we divide the player foreground region into 16 *contour segments*, centered on the player centroid. Over the entire stroke, we extract contour features for each video frame. We then normalise the features in order to make them invariant to the player's distance from the camera. A binary Bayesian classifier was trained on samples of left handed players serving and right handed players serving. At the beginning of each match, the first five serves from each player are used as input vectors and the classifier identifies in which hand the player is holding the racquet.

3.1.6 Rally and Game Detection

The occurrence of a rally can be inferred by detecting all the non-serve strokes which occur between two serves. The start of a new game can be inferred from serve detection as we recognise when the serve switches from Player A to Player B and vice versa. Tie-break games are easily detected using a rule-based algorithm since the serve direction will change after the first point in a tie break game and every two points thereafter.

(a) Camera image (b) Foreground (c) Radial map

Figure 2: Contour feature extraction.

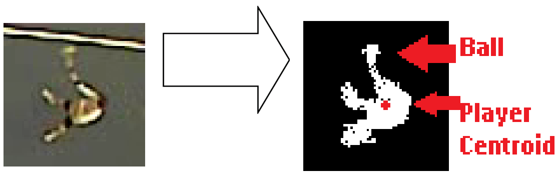

Figure 3: Player centroid and ball location are compared on the y-axis to determine if the ball is struck above or below a player

3.2 Inertial Event Indexing

In this section, we give a brief overview of our tennis stroke detection system using WIMUs, which has been previously published [2]. Strokes are detected by fusing the accelerometer, gyroscope and magnetometer sensors at an early stage. A WIMU sensor placed on a player's dominant arm events are classified from the sensor data of a single sensor.

3.2.1 Stroke Classification

We classify the main types of tennis strokes (forehand, backhand and serves) played in a competitive match. A two-level Bayesian classification system is used, whereby the first step filters any player movements where they are not performing a tennis stroke and the second step uses a Bayesian Network to classify the remaining candidate tennis strokes into serves, backhands or forehands [2].

3.2.2 Inferring Rallies, Games & Change of Ends

To find all rally boundaries, all forehands and backhands are grouped until a new serve occurs. To detect game boundaries all rallies are grouped together until the serve changes from one player to the next. Since players change ends at the end of the first and every subsequent two games as defined by ITF, we can use the game boundaries to infer a new change of end event.

4. SENSOR DATA INTEGRATION

Event indexing is first executed from separate sensors and then both sensors are synchronised, before selected events from each sensor are imported into a relational database. This indexing system may work in real time in the future, but for now data integration is performed offline. The visually detected events which are imported are player and ball tracking and change of end events. The inertial sensors provide forehand, backhand, serve events for both players. SQL integration queries then map all player strokes in the DB to the relevant player and ball tracks using a rule-based query. Each high-level event in the following section is detected offline, which allows users to retrieve complex high-level queries in a matter of seconds.

4.1 High-level Query Generation

Having detected the strokes played (inertial sensors), along with the player and ball tracks (visual sensors), we can detect the movement of the ball after a specific stroke is executed by a given player.

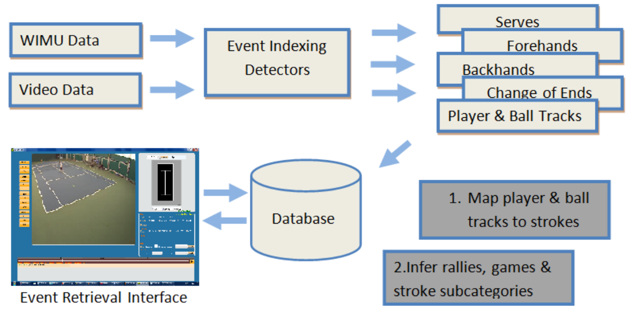

Figure 4: Event Retrieval & Detection Overview

An offline rule-based query then detects if the serve is a **T, Body or Wide** and a similar rule-based query detects if a backhand or forehand is **in-to-in, in-to-out, cross, line**.

A rule based query engine is executed offline to identify **first serves made/missed**. This module uses heuristics based on the server's movements immediately after a serve. The rules of tennis state that if the first serve is illegal, the second serve must be taken from the same side of the baseline. Therefore if a serve is executed and the servers location remains on the same side of the baseline for the next serve, then we can infer that the previous serve was missed, otherwise the first serve was made. Following on from the knowledge of first serves made/missed, the strokes played by the returning player are analysed offline to infer the **return of first serves** and **return of second serves**. A rule-based query identifies which player is returning and what type of stroke is being executed.

Ball tracks are used to detect which strokes result in the ball hitting the **net**. This query, which is again executed offline, simply records the origin of a each stroke played. If the ball track terminates in what is defined as the net region, it is assumed that the ball has hit the net.

A **volley** is usually played when a player returns the ball, while positioned close to the net, hence the ball does not have time to hit the ground and is volleyed. To detect forehand and backhand volleys, a query detects forehands and backhands with inertial sensors and then cross examines each detected stroke with the relevant player track to infer if the player is within volley range of the net. A **smash** stroke will have a player executing the same motion as a serve except the smash will be played during a rally, unlike a serve which is always the first stroke in a rally. To detect a smash a rule-based query analyses the inertial stroke recognition data and labels any strokes which are classified as a serve, but occur while the player is positioned inside the baseline as a smash.

5. EVENT RETRIEVAL SYSTEM

The retrieval interface (Figure 5) consists of four main panels: the player panel displays the names of the players indexed, the match time line displays all the matches played by a selected player from the player panel. Each time line represents a single match.

5.1 Events Panel

The events panel provides an interface for users to build specific queries based on the high-level events determined in Section 4.1. For example, from our coaches we know that the user might want to view the video of all the T-Body serves executed by *Player A* from the left side of the baseline. Each event retrieved from the match will then be represented along the match timeline as a vertical tick, which can be played by clicking on the tick.

5.2 User Query Panel

This panel allows users to visually construct a query by drawing rectangles and line objects which represent players and a ball in flight. The user can then retrieve all strokes which are played while both players are simultaneously inside their respective rectangle and the ball travels along the drawn shot line. Users can also filter stroke types and the results are displayed along the match timeline for efficient and visual retrieval.

A rule-based query firstly filters out the relevant strokes of a requested player. Then the player tracks find the temporal locations where a given stroke type is executed whilst both players are inside the localised coordinates drawn by a user. Finally the flight of the ball is inferred by converting the ball flight line drawn by the user to real world coordinates.

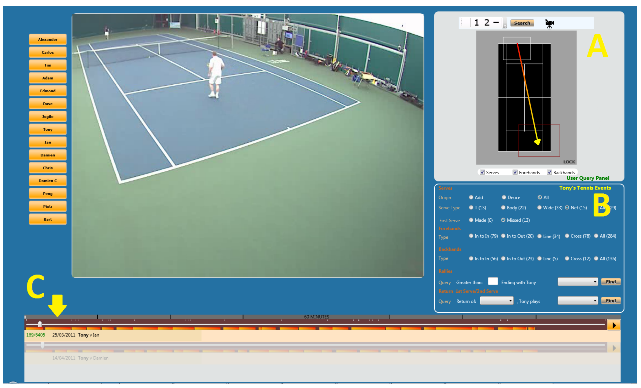

Figure 5: Event Retrieval Interface, (A) User Query Panel, (B) Event panel (C) Match timeline panel is used to display & playback events.

6. EXPERIMENTS

6.1 Experiment One

The first experiment is to asses how accurate event detection is using inertial sensors compared to visual sensors. For this experiment we classify a number of events. These are: serves, backhands, forehands, dominant arm detector (video only), rallies, games and change of end events. 12 complete tennis matches (825 minutes) from various skill levels were captured and a ground truth was generated. If the type of the detected stroke is the same as the ground truth and the impact time of a serve, backhand or forehand is ± 1 second, we counted it as a true detection. The impact times for rallies and games were ± 2 and ± 10 respectively. For video, we apply the dominant arm detector to all players. The precision and recall results for video are .84 and .85 respectively. For inertial sensors the precision and recall results are .95 and .85 respectively.

6.2 Experiment Two

There is no published work which detects strokes using inertial sensors, therefore to compare our approach to [6], we selected the same six events as reported in [6] to be recognised: forehand, backhand, service, smash, forehand volley and backhand volley. 240 sequences were selected, which are performed by 5 players during real matches. Inertial sensors are used to detect serves, forehands and backhands. Forehand volley, backhand volley and smashes were detected as described in Section 4.1. Stroke recognition classifiers were trained with one group of players and the evaluation set contained strokes from unseen players. Although not directly comparable to [6], our recognition results for this experiment are 82% which is in line with the results found in [6]. Furthermore, our dataset contains novice players, unlike in [6] where only professional players from broadcast video are used. This proves that our system can work on players of all levels and this is necessary for a coaching tool. This stroke recognition system was evaluated on an Intel Core 2 Duo Processor and the combined training and recognition time of a single event took on average 7 seconds.

6.3 Experiment Three

In this experiment we test the accuracy of retrieving high-level tennis events from the user interface. From a ground truth of 5 different matches, we selected a random 10 instances of each of the following: serves (T, Body or Wide), forehands (in-to-in, in-to-out, cross, line), backhands (cross, line), first serve made/missed, return of first serve/second serve and backhand net. The retrieval accuracy was 74%.

7. CONCLUSIONS AND FUTURE WORK

We conclude that inertial sensors do perform better than visual sensors at event detection in tennis. However visual sensing is non-intrusive and the overhead camera gives player and ball locations, which cannot be determined from inertial sensors alone. The visual and inertial event indexing component can infer stroke sub categories (T serve, forehand in to in etc) which have not been automatically indexed before and it would take manual editors an inordinate amount of time to achieve this level of event indexing.

Future work for this research is to automatically track player and ball movements from a single view-invariant camera, making it possible to use this system in tournaments. We currently do not detect set boundaries or scores, but this is targeted for future research. Investigations are also required to replace the WIMUs devices with a mobile phone, since smartphones nowadays contain accelerometers and gyroscopes.

8. ACKNOWLEDGMENTS

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

9. REFERENCES

[1] D. Conaghan, P. Kelly, and N. E. O'Connor. Game, shot and match: Event-based indexing of tennis. In *9th International Workshop on Content-Based Multimedia Indexing*, pages 96–102, 2011.

[2] D. Connaghan, M. Gaffney, P. Kelly, M. Walsh, C. O'Mathuna, and N. E. O'Connor. Multi-sensor classification of tennis strokes. In *Proceedings of the IEEE SENSORS 2011 Conference*, 2011.

[3] M. Delakis, G. Gravier, and P. Gros. Audiovisual integration with segment models for tennis video parsing. *Comput. Vis. Image Underst.*, 111:142–154, August 2008.

[4] M. Gaffney and B. O'Flynn. Wearable wireless inertial measurement for sports applications. In *IMAPS-CPMT Poland Conference*, 2009.

[5] K. Ke, T. Zhao, and O. Li. Bhattacharyya distance for blind image steganalysis. *Multimedia Information Networking and Security, International Conference on*, 0:658–661, 2010.

[6] M. Petkovic, W. Jonker, and Z. Zivkovic. Recognizing strokes in tennis videos using hidden markov models. In *VIIP*, pages 512–516, 2001.